

INCORPORATING ASTROPHYSICAL SYSTEMATICS INTO A GENERALIZED LIKELIHOOD FOR COSMOLOGY WITH TYPE IA SUPERNOVAE

KARA A. PONDER*, W. MICHAEL WOOD-VASEY, ANDREW R. ZENTNER
 Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC).
 Physics and Astronomy Department, University of Pittsburgh, Pittsburgh PA, 15260, USA
Draft version October 17, 2016

ABSTRACT

Traditional cosmological inference using Type Ia supernovae (SNeIa) have used stretch- and color-corrected fits of SN Ia light curves and assumed a resulting fiducial mean and symmetric intrinsic dispersion for the resulting relative luminosity. As systematics become the main contributors to the error budget, it has become imperative to expand supernova cosmology analyses to include a more general likelihood to model systematics to remove biases with losses in precision. To illustrate an example likelihood analysis, we use a simple model of two populations with a relative luminosity shift, independent intrinsic dispersions, and linear redshift evolution of the relative fraction of each population. Treating observationally viable two-population mock data using a one-population model results in an inferred dark energy equation of state parameter w that is biased by roughly 2 times its statistical error for a sample of $N \gtrsim 2500$ SNeIa. Modeling the two-population data with a two-population model removes this bias at a cost of an approximately $\sim 20\%$ increase in the statistical constraint on w . These significant biases can be realized even if the support for two underlying SNeIa populations, in the form of model selection criteria, is inconclusive. With the current observationally-estimated difference in the two proposed populations, a sample of $N \gtrsim 10,000$ SNeIa is necessary to yield conclusive evidence of two populations.

Subject headings: supernovae: general, cosmological parameters, methods: statistical

1. INTRODUCTION

Type Ia supernovae (SNeIa) are excellent standardizable candles that enabled the discovery of the expansion of the universe in the late 1990s by Riess et al. (1998) and Perlmutter et al. (1999). Originally, SNeIa were used as standard candles from empirical evidence with a scatter of only ~ 0.3 magnitudes (Baade 1938; Kowal 1968). As data sets grew, patterns appeared in the light curves yielding the brighter-slower (Phillips 1993) and brighter-bluer (Riess et al. 1996; Tripp 1998) relationships, which standardized the supernovae further by reducing their scatter down to ~ 0.15 magnitudes.

The goal of this paper is to create a framework to properly model effects that change the *distribution* of expected SN Ia apparent brightness at each redshift. If unmodeled, these effects lead to systematic biases in cosmological inference. We propose using general and flexible likelihood functions that have the ability to handle insufficiently modeled systematics. As an example, we simulate a simplistic toy model of two SN Ia populations with a small relative shift in absolute magnitude. The relative rate of these two populations changes linearly with redshift. We examine the systematic errors in cosmological parameters caused by incorrectly fitting multiple populations with a single Gaussian model and show that these errors can be eliminated by using a multiple population model to fit the SN Ia magnitude-redshift relation. In this paper, we focus on this toy model to demonstrate the validity of this framework. The consideration of more complex multiple-population models or other astrophysical or observational effects that lead to shifting magnitude distributions with redshift will be

considered in subsequent papers.

Though the two population model is intended as an example, there are several motivators for multiple populations of SNeIa. For instance, after adjusting the light curves with these observed relationships, there is still an unaccounted for feature in the corrected brightness residual with respect to the distance-redshift relationship (Hubble residual) that appears to be correlated with host galaxy properties. In the last five years, there have been myriad studies (Kelly et al. 2010; Sullivan et al. 2010; Lampeitl et al. 2010; Gupta et al. 2011; Johansson et al. 2013; Childress et al. 2013; Rigault et al. 2013, 2015; Kelly et al. 2015) comparing host galaxy mass, metallicity, and/or star formation rate to residuals in the Hubble diagram. Rigault et al. (2013) examined the relationship between global and local star formation rates through H-alpha and found that SNeIa in locally passive environments were brighter than those in locally star forming environments. Rigault et al. (2015) and Kelly et al. (2015) used *GALEX* ultraviolet observations and confirmed this correlation between Hubble residual and local star-formation rate.

It is possible that the host galaxy correlations are caused by something more fundamental such as the nature of the progenitor. Though the evidence for host galaxy correlations may be controversial (Jones et al. 2015), there is increasing evidence that there are two different progenitor channels that could create a multiple population effect (Greggio 2005; Cao et al. 2015; Olling et al. 2015).

The most recent analysis of SNeIa for cosmology comes from Betoule et al. (2014) with the Joint Lightcurve Analysis (JLA). They account for the observed correlation between Hubble residual and host galaxy mass by

* Email: kap146@pitt.edu

creating a step function for the absolute magnitude of each supernova based on the host galaxy mass. They then implicitly assume a Gaussian likelihood and fit for parameters using a χ^2 method. We will expand this method by defining a continuous function for absolute magnitude and finding the most probable parameter regions with a generalized likelihood through Markov Chain Monte Carlo (MCMC) techniques.

Being able to identify and accurately correct for systematics is becoming more important as the number of SNeIa drastically increases with current surveys such as Dark Energy Survey (DES)¹, Panoramic Survey Telescope and Rapid Response System (Pan-STARRS Scolnic et al. 2014; Rest et al. 2014). The amount of SN Ia data available for cosmological analyses will continue to increase into the future with surveys such as the Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al. (2009)), Wide-Field Infrared Survey Telescope-Astrophysics Focused Telescope Asset (WFIRST-AFTA, Spergel et al. (2015)), and the European Space Agency’s Euclid² mission on the horizon (Astier et al. 2014). Supernova cosmology is no longer statistically limited and is rapidly becoming systematically limited. Now is the time to explore different avenues for undertaking unbiased cosmological analyses with large data sets.

In Section 2 we discuss non-Gaussian error distributions as modeled by multiple Gaussian populations. Section 3 defines how mock SN Ia data sets are generated. Here, we introduce a toy model that represents a redshift evolution of the populations to probe uses of the framework. In Section 4 we define the likelihood to be used in the MCMC and the different model selection techniques. Section 5 shows that both population and cosmological parameters are biased if multiple populations are not included in the analysis. Though it has more model parameters, the Gaussian mixture model recovers input cosmology with only a $\sim 1-3\%$ loss in precision. We show that current and future data sets will have the cosmology biased before there are enough statistics to characterize the underlying systematic or to robustly require a more complicated model. Section 6 discusses how the models presented here relate to current cosmological analyses and presents possible astrophysical motivations for multiple populations. In Section 7 we summarize our results and discuss ways to improve and expand this framework.

2. NON-GAUSSIAN ERROR DISTRIBUTIONS

The most commonly used method for cosmological parameter estimation in supernova cosmology is χ^2 minimization. Implicit in this method is the assumption that the overall probability density function (PDF) of supernovae follows a Gaussian distribution or can be linearly corrected to do so. With motivations such as the host galaxy correlations, complexities in the analysis from Malmquist bias, and uncertainties about dust, there are too many unknowns within supernova cosmology for SNeIa to be adequately described by a single point estimator in a Gaussian PDF. The PDF of SNeIa needs to be expanded to be able to more fully model the systematic

effects underlying the observed luminosity distribution of SNeIa.

Here we will explore one possible expansion to the PDF of SNIa luminosity. While this example is inspired by the recent discussions of correlations between SNIa corrected luminosity and host galaxy properties, its use here is intended as a demonstration of the framework. We are not arguing for any particular specific model as being representative of the SNIa population extant in the Universe.

2.1. Gaussian Mixture Models

Karl Pearson popularized using multiple Gaussians to describe non-Gaussian data in 1894 when he showed that two Gaussians were a better fit to crab morphologies which strengthened the claim for evolution (Pearson 1894).³

A distribution consisting of multiple Gaussian populations with different peaks and/or dispersions is referred to as a Gaussian mixture model (GMM) and the probability density function (PDF) that describes it is

$$p_{\text{GMM}}(x) = \sum_{j=1}^{\mathcal{N}} \frac{n_j}{\sqrt{2\pi\sigma_j^2}} \exp\left(\frac{-(x-\lambda_j)^2}{2\sigma_j^2}\right) \quad (1)$$

where \mathcal{N} is the number of populations; and for each population j : n_j is the relative normalization ($\sum_{j=1}^{\mathcal{N}} n_j = 1$); λ_j is the mean; and σ_j is the standard deviation.

For the sake of simplicity and because it is motivated by current observational literature, in this paper we focus on a model with only two populations: A and B. Under this model Eq. 1 then becomes

$$p_{\text{GMM}}(x) = \frac{n_A}{\sqrt{2\pi\sigma_A^2}} \exp\left(\frac{-(x-\lambda_A)^2}{2\sigma_A^2}\right) + \frac{n_B}{\sqrt{2\pi\sigma_B^2}} \exp\left(\frac{-(x-\lambda_B)^2}{2\sigma_B^2}\right). \quad (2)$$

There are five parameters that need to be specified: λ_A , λ_B , σ_A , σ_B , and n_A (n_B is implicitly specified under the constraint that $n_A + n_B = 1$). Once the PDF has been defined, the log-likelihood function for the two-population model, \mathcal{L} , is simply

$$\mathcal{L} = \ln L = \sum_{i=1}^N \ln \left[\frac{n_A}{\sqrt{2\pi\sigma_A^2}} \exp\left(\frac{-(x_i-\lambda_A)^2}{2\sigma_A^2}\right) + \frac{n_B}{\sqrt{2\pi\sigma_B^2}} \exp\left(\frac{-(x_i-\lambda_B)^2}{2\sigma_B^2}\right) \right], \quad (3)$$

where N is the total number of objects included in the analysis, x_i is some observed quantity per object, and (λ_A, σ_A) , (λ_B, σ_B) are the model mean and standard deviation for the populations A and B.

In the case of SN Ia cosmology, x_i is the observed width-color-corrected apparent magnitude of supernovae, and (λ_A, σ_A) , (λ_B, σ_B) would correspond to

¹ <http://www.darkenergysurvey.org/>

² <http://sci.esa.int/euclid/>

³ Thanks to S. Peng Oh for this reference.

models of two different SN Ia populations with different absolute magnitudes and intrinsic dispersions, each propagated through the same cosmological model for the luminosity distance modulus.

3. GENERATING MOCK DATA SETS

We begin exploring a two-population GMM for SNeIa by generating a sample of mock SN Ia data sets from Eq. 2. We represent the difference in the two populations as a difference in absolute magnitude M_X for $X = A$ or B populations. The parameters in Eq. 2 can thus be re-defined as: $\lambda_A \rightarrow M_A$, $\lambda_B \rightarrow M_B$. While we will discuss absolute magnitude distributions in this section in order to emphasize the different populations, later we will consider fitting the mock data as “observed” apparent magnitudes. We define the relative mean magnitude shift between the populations such that $\Delta M \equiv M_A - M_B$ and re-parameterize M_B in terms of M_A and ΔM as $M_B = M_A - \Delta M$. The relative magnitude difference ΔM is thus applicable to either absolute or apparent magnitude, and the overall normalization of the absolute magnitude – which is generally marginalized over – is absorbed into one term for both populations. The variance of each population (σ_X^2) is defined as $\sigma_X^2 = \sigma_{\text{int},X}^2 + \sigma_{\text{phot}}^2$ including the intrinsic dispersion of the population $\sigma_{\text{int},X}$ and the dispersion introduced from observational errors σ_{phot} .

Figure 1 illustrates graphically the five parameters of our two-population GMM: M_A , ΔM , σ_A , σ_B , and n_A and the two parameters of a single-Gaussian model (SGM): M and σ fit to the GMM-generated data. For visual clarity, this example has $n_A = 0.7$ and shows an extreme shift of $\Delta M = 1.0$ mag. We expect realistic models to be on the order of $\Delta M \lesssim 0.1$ mag.

We simulate mock data sets assuming the peaks of the populations average to the estimated value of M such that $(M_A + M_B)/2 = -19.5$ mag with intrinsic dispersions of $\sigma_{\text{int},X} = 0.1$ mag and $\sigma_{\text{phot}} = 0.1$ mag for both populations. σ_{phot} was chosen to reflect the observational error that JLA achieved (~ 0.115 mag). The supernovae are constrained to a redshift range of $0.05 < z < 1.5$ to cover the low redshift anchors and the high redshift cosmology probes.

Because host galaxy properties are on average different at $z \sim 0$ and $z \sim 1$, it becomes sensible to explore the possibility of redshift evolution between the relative number of SNeIa in each population. As a toy model we simulate a redshift dependence of the relative normalizations by having the populations evolve linearly in redshift: $n_A = n'_{A,0}z + n_{A,0}$. Where $n_{A,0}$ is $n_A(z)$ evaluated at $z = 0$ and $n'_{A,0}$ is the first derivative of $n_A(z)$ evaluated at $z = 0$. We then impose boundary conditions such that the total population of supernova is dominated by a single population at the lowest redshift $n_A(z_{\text{min}} = 0.05) = 1$ and the other population dominates the total population at the highest redshift $n_A(z_{\text{max}} = 1.5) = 0$ to get $n_{A,0} = 1.003$ (no units) and $n'_{A,0} = -0.627$ in units of $1/\text{redshift}$. The two populations have an equal number of supernovae at $z = 0.775$ as set by the slope and intercept of $n_A(z)$. This value is derived only from relative normalizations and is independent of other supernova population parameters.

A linear evolution with redshift is an overly simplistic

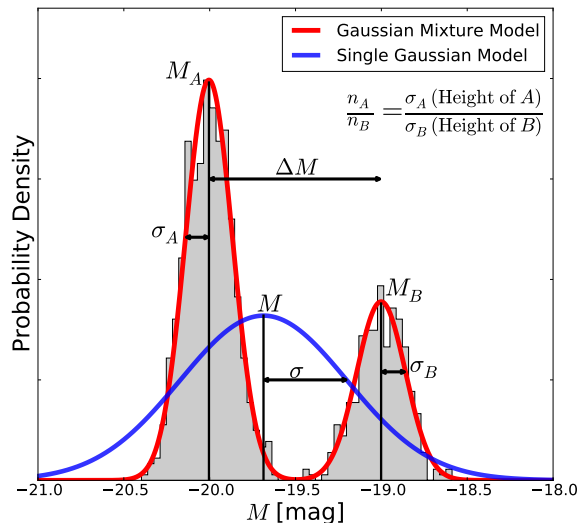


Figure 1. A histogram of mock supernovae with two populations are shown in grey and fit with a GMM and SGM. M and σ are the peak and dispersion from a SGM. The GMM model includes the location of both peaks, M_A and M_B , and the dispersion of both populations, σ_A and σ_B . ΔM is the difference between M_A and M_B in magnitudes. The relative number of SNeIa in each population is n_A/n_B (where $n_A + n_B = 1$). In this example, $\Delta M = 1.0$ mag and $n_A = 0.7$.

model. The evolution of multiple populations or other astrophysical systematics will likely be a smooth, potentially monotonic, function of redshift. While a power law or logarithmic function might suggest itself as a good model for a variety of phenomena, a linear dependence is at least a reasonable description of a function for which we have a strong bias that should be varying slowly. As such, it is informative to explore a linear model, which is likely to capture a significant amount of the overall trend of the true astrophysical systematics. In Greggio et al. (2008), Figure 7 (top panel) shows the relative rates of the single degenerate channel versus double degenerate channels as a function of redshift. These are clear parallels to our relative population parameters, and one of the models shows a linear trend. The modeling of SNeIa progenitors is still incomplete and different models can provide drastically different rates. The GMM does not rely on a linear model for the evolution of the relative populations and can easily be constructed with different forms such as a power law or logarithmic function.

We randomly draw a redshift from a uniform distribution in the range $0.05 < z < 1.5$, then generate a GMM PDF corresponding to that redshift, and randomly draw an absolute magnitude from that PDF.

Figure 2 illustrates how the absolute magnitude distribution of SNeIa evolves with redshift for two different ΔM s. While the redshift evolution is a small effect for small ΔM , the shift between different populations becomes understandably more clear when $\Delta M = 0.5$ mag.

We generate 108 different data sets with a range of number of supernovae in each set: $N = 100, 1000, 2500, 10000$ and a range of shifts between the two supernovae populations: $\Delta M = 0.0, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ mag. Each

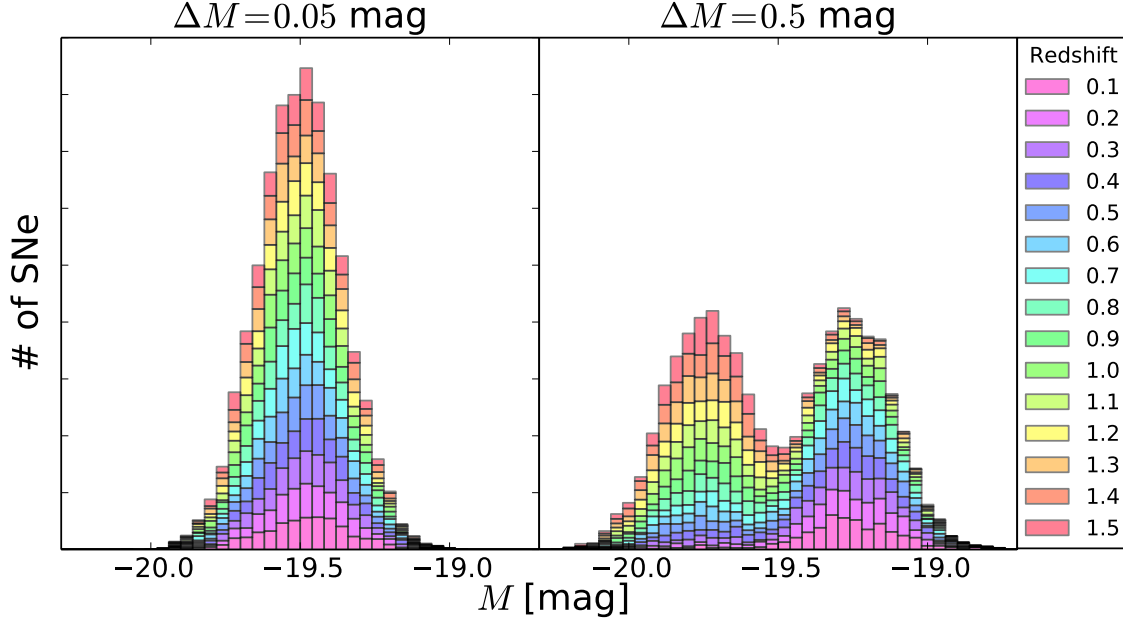


Figure 2. Absolute magnitude distribution of 10,000 mock SNeIa separated into 15 redshift bins denoted by color. A histogram is generated at each redshift then stacked upon the previous redshift’s histogram. *Left:* A small separation of $\Delta M = 0.05$ mag is a subtle shift. *Right:* A exaggerated separation of $\Delta M = 0.5$ mag makes the evolution visually obvious.

permutation of N with ΔM is performed three times to help average over random fluctuations in the data sets. The number of supernovae correspond to a small sample, the order of current data sets (1000), and the expected yields from WFIRST-AFTA (2500) and LSST (10,000).⁴ $\Delta M = 0.0$ mag is consistent with a single Gaussian population while $\Delta M = 0.1$ mag is close to the number quoted from Rigault et al. (2015) for the difference in brightness between supernovae located in active versus passive local environments. Though this framework is discussed with a specific systematic as an example motivation, it is general and can be applied to any systematic that can be described by an effective distribution in the likelihood.

In order to use apparent magnitudes instead of absolute magnitude, we add the cosmological distance modulus $\mu(z; \Omega_M, w)$ to produce an apparent magnitude (m). We chose our default cosmology to be that of WMAP9 with $\Omega_M = 0.2865$, $\Omega_L = 0.7134$, $w = -1$, $H_0 = 69.32$ km Mpc⁻¹ s⁻¹ (Hinshaw et al. 2013). We do not simulate a distribution of stretch and color or the resulting correction process. This process is thus rather generically applicable to any luminosity distance indicator with no particular restriction to SNeIa beyond the parameters chosen for the GMM.

In the present work, we also neglect the effects of gravitational lensing on SN Ia analyses. Though the dispersion induced by lensing may be non-negligible in forthcoming analyses (Zentner & Bhattacharya 2009), lensing does not shift the average brightness (setting aside observational selection effects for the moment) and is unlikely to bias cosmological results (Helbig 2015). We defer a more complex analysis including lensing to future work.

⁴ Current estimates of cosmologically useful SNeIa from LSST range from 10,000s to 100,000. We have chosen a very conservative value here.

4. METHODS

4.1. Markov Chain Monte Carlo

We use standard Markov Chain Monte Carlo (MCMC; Metropolis et al. 1953) techniques to fit for model parameters. In particular, we utilize the affine-invariant ensemble sampler from Goodman & Weare (2010) and implemented in `python` with `emcee` (Foreman-Mackey et al. 2013). We test the convergence of our chains by checking that the autocorrelation of points sampled from the posterior approaches zero for large lags (Box & Jenkins 1976).

The likelihood including cosmology used for the MCMC analysis is defined as

$$\mathcal{L} = \sum_{i=1}^N \ln \left[\frac{n_A(z)}{\sqrt{2\pi\sigma_A^2}} \exp \left(\frac{-(m_i - m_A)^2}{2\sigma_A^2} \right) + \frac{(1 - n_A(z))}{\sqrt{2\pi\sigma_B^2}} \exp \left(\frac{-(m_i - m_B)^2}{2\sigma_B^2} \right) \right] \quad (4)$$

where:

- N is the number of supernovae in the mock data set;
- $n_A(z)$ is the relative normalization of population A,

$$n_A(z) = n'_{A,0}z + n_{A,0} \quad ;$$

- σ_X is the standard deviation of the two populations such that

$$\sigma_X^2 = \sigma_{\text{phot}}^2 + \sigma_{\text{int},X}^2 \quad \text{where } X = A \text{ or } B;$$

- m_i is the generated “observed” apparent magnitude for supernova i in the mock data set;

Table 1
Flat Priors

Ω_M	w	$\mathcal{M}/\mathcal{M}_A$	ΔM	$\sigma_{\text{int},X}$	$n'_{A,0}$	$n_{A,0}$
[0,1]	[-3,1]	[-10, 5]	[0, 5]	[0.0, 0.3]	[-1,0]	[0, 2]

- m_A and m_B are predicted apparent magnitudes based on cosmological parameters through the Hubble constant-free luminosity distance,

$$\begin{aligned}
m_A &= 5 \log(\mathcal{D}_L(z; \Omega_M, w)) + \mathcal{M}_A \\
&\quad \text{where } \mathcal{M}_A = 25 - 5 \log H_0 + M_A \\
&\quad \text{and} \\
m_B &= 5 \log(\mathcal{D}_L(z; \Omega_M, w)) + \mathcal{M}_B \\
&\quad \text{where } \mathcal{M}_B = \mathcal{M}_A - \Delta M.
\end{aligned}$$

We assume a flat universe ($\Omega_M + \Omega_\Lambda = 1$) and fit for the matter density Ω_M and the dark energy equation of state parameter w . In the case of the GMM fits, we also fit for six nuisance parameters: \mathcal{M}_A , ΔM , $\sigma_{\text{int},A}$, $\sigma_{\text{int},B}$, $n'_{A,0}$ and $n_{A,0}$ which incapsulate the information about the underlying SN Ia populations. However, since we used the Hubble constant free luminosity distance, we must still specify H_0 to completely describe the underlying populations.

In addition to our GMM analysis, we also fit each data set using a single-Gaussian model (SGM) for the underlying SN Ia population; these fits have just two nuisance parameters: \mathcal{M} and σ_{int} .

For all parameters we use the flat priors defined in Table 1 and an extra prior in the GMM on the combination of $n'_{A,0}$ and $n_{A,0}$ such that $0 \leq n_A(z) \leq 1$.

4.2. Model Comparison

We have introduced a GMM to treat the cosmological analyses of SN Ia data. The GMM is more complex than the SGM as evidenced, in part, by the fact that the GMM has four more nuisance parameters. The question arises whether or not the additional complexity is demanded by the data or, in our case, by the mock data used to mimic forthcoming analyses. We employ three statistical tests to indicate whether or not the additional complexity is required by the data: the Akaike Information Criterion (AIC; Akaike 1974); the Bayesian or Schwartz Information Criterion (BIC; Schwarz 1978); and the Deviance Information Criterion (DIC; Spiegelhalter et al. 2002). For a review of these three methods we refer the interested reader to Liddle (2007) and for a more in-depth discussion of AIC and DIC see Gelman et al. (2014).

The AIC and BIC are calculated from the maximum likelihood \mathcal{L}_{max} , the number of model parameters k , and the number of data points N as

$$\text{AIC} = -2 \ln \mathcal{L}_{\text{max}} + 2k + \frac{2k(k+1)}{N-k-1} \quad (5)$$

and

$$\text{BIC} = -2 \ln \mathcal{L}_{\text{max}} + k \ln N. \quad (6)$$

Models with lower values of these information criteria are favored. Both the AIC and BIC penalize models

with a greater number of parameters (greater k) because \mathcal{L}_{max} can only increase with increased parameter freedom, while the BIC also penalizes larger data sets (greater N) to reduce the risk of over fitting.

The DIC is more suited for analyses with MCMC outputs because it directly uses the resulting samples from the posterior. The DIC can be computed from these samples in the MCMC chain as

$$\text{DIC} = 2\overline{D(\boldsymbol{\theta})} - D(\tilde{\boldsymbol{\theta}}), \quad (7)$$

where $\boldsymbol{\theta}$ is the set of parameters directly from the samples in the chain (in our case these are the cosmological parameters w and Ω_M along with the parameters of either the SGM or GMM), $D(\boldsymbol{\theta})$ is the deviance,

$$D(\boldsymbol{\theta}) = -2 \ln \mathcal{L}(\boldsymbol{\theta}) + C, \quad (8)$$

$\mathcal{L}(\boldsymbol{\theta})$ is the likelihood evaluated at parameters $\boldsymbol{\theta}$, and C is a normalizing constant that cancels when comparing different models. $\overline{D(\boldsymbol{\theta})}$ is the average of the deviance evaluated at each step in the chain and $D(\tilde{\boldsymbol{\theta}})$ is the deviance evaluated at the mean, median, or some other summary point in parameter space $\tilde{\boldsymbol{\theta}}$. In our samples, we find that the median is a better representation of our data because many of the posterior distributions are non-Gaussian, which can result in a mean value strongly influenced by tails.

5. RESULTS

5.1. An Illustration of Parameter Bias

We illustrate the potential for bias in the inferred cosmological parameters due to multiple SN Ia populations by first presenting Hubble diagrams. We consider data generated from an underlying GMM but fit with both a SGM likelihood and a GMM likelihood. The fit using a SGM likelihood function is intended to mimic an analysis in which there is no mechanism to account for two distinct populations.

Figure 3 shows the results of a comparison between a SGM and GMM analysis using one data set with an exaggerated shift in the magnitude difference between the two populations, $\Delta M = 0.5$ mag. We use this large shift here for illustrative purposes and more realistic values are $\Delta M \lesssim 0.1$. The upper panel of the left figure in Figure 3 shows, within ten evenly-spaced redshift bins, the PDF of apparent magnitude inferred from both the SGM and GMM fits to the underlying, multi-modal, GMM mock data. The parameters of these PDFs are determined by the fits through the MCMC process described in Section 4 with the cosmological parameters held constant for simplicity. The SGM was fit at each redshift bin while the GMM was fit using all the data at once to constrain the parameters of redshift evolution. The peak of the SGM PDF in the residual ($m_{\text{data}} - m_{\Lambda\text{CDM}}$) exhibits a linear evolution getting brighter as redshift increases, which is the result of the redshift evolution in the data set.

The right plot in Figure 3 shows the same data set and MCMC fit (with cosmology constant) converted into distance modulus versus redshift. Simply subtracting the absolute magnitude derived from the MCMC fit of the mock data yields this information. The absolute magnitude for the SGM can be taken straight from the chains

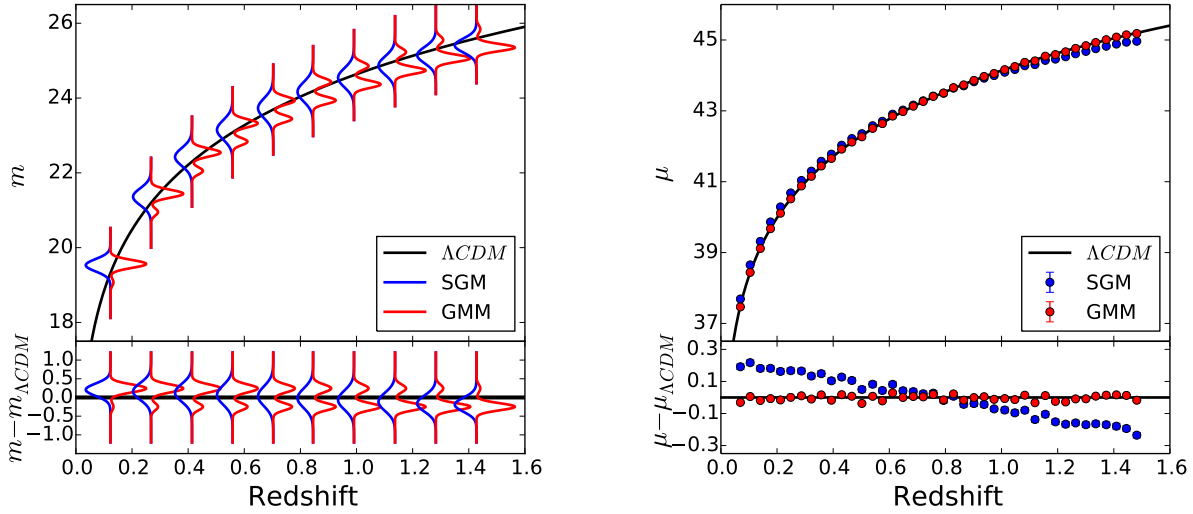


Figure 3. *Left:* Apparent magnitude versus redshift (top panel) and the Hubble residual (bottom panel) for parameter inferences using a GMM with $\Delta M = 0.5$ mag and $N = 10,000$ SNeIa. The black line cutting across the data is the expected magnitude redshift relation in our fiducial Λ CDM cosmology. At each of ten evenly-spaced redshift bins the blue (left-directed) curves show the PDFs of m inferred from a SGM fit to the GMM mock data while the red (right-directed) curves show the PDFs of m inferred by fitting the data with a GMM. Clearly the inferred m are biased in the SGM fits to the GMM mock data. *Right:* Luminosity distance modulus versus redshift (top panel) and residual (bottom panel) for $\Delta M = 0.5$ mag with $N = 10,000$. The line is the distance modulus calculated from Λ CDM. The data points are the mock data sets minus the model for absolute magnitude ($\mu = m - M_{\text{model}}$) using the SGM (blue) with $M_{\text{model}} \equiv M$ and the GMM (red) with $M_{\text{model}} \equiv n_A M_A + n_B M_B$ with values derived from fit models holding cosmology and H_0 constant. 1σ error bars have been plotted but are too small to see.

($M_{\text{SGM}} = M$); however, the absolute magnitude for the GMM is a function of redshift and multiple fitted parameters ($M_{\text{GMM}}(z) = n_A(z)M_A + n_B(z)M_B$). The inferred SN Ia population parameters M and σ for the SGM have no way to account for the relative shift between the two SN Ia populations as a function of redshift and so the SGM fits show a systematic, redshift-dependent deviation in the distance modulus as a function of redshift. Notice that the mock GMM data set was generated such that at $z = 0.775$, the two populations have an equal number of SNeIa and, as expected, $M_{\text{SGM}} = M_{\text{GMM}}$ at $z = 0.775$. The population parameters are recovered well for the GMM fit and there is clearly no bias in this case.

5.2. Cosmological Parameters

From the perspective of exploiting SNeIa as a probe of cosmology, the greatest concern caused by multiple populations of SNeIa is that insufficiently accurate modeling of the multiple populations will lead to biased cosmological parameters. Exploring this possibility is the primary purpose of this paper. To explore the potential importance of multiple SN Ia populations on cosmology, we fit each of the 108 mock data sets described in Section 3 for the cosmological parameters, Ω_M and w , and SN Ia population parameters simultaneously.

Figure 4 displays the Hubble diagram inferred from both SGM and GMM fits to a GMM model from a single data set with $N = 10,000$ SNeIa and an extreme value of $\Delta M = 0.5$ mag. This large value of ΔM is used to produce this figure only because it has the pedagogical value of making the influence of the two-populations model on inferred cosmology obvious. Clearly the GMM fits yield an unbiased Hubble diagram and we infer unbiased values of both Ω_M and w .

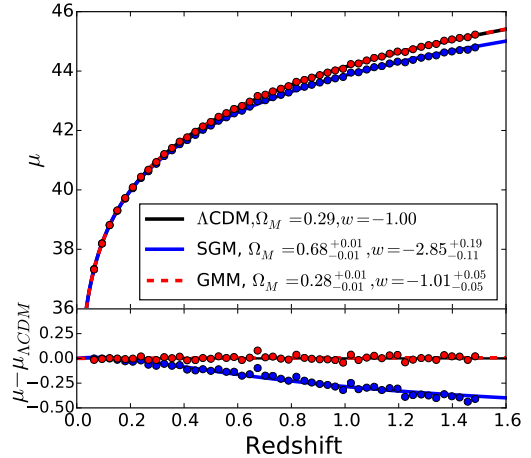


Figure 4. Hubble constant-free luminosity distance modulus versus redshift (top panel) and residual (bottom panel) for $\Delta M = 0.5$ mag with $N = 10,000$. The lines are the distance modulus calculated from cosmology derived from the MCMC fits. The data points are the mock data sets minus the model for absolute magnitude ($\mu = m - M_{\text{model}}$) using the SGM (blue) with $M_{\text{model}} \equiv M$ and the GMM (red) with $M_{\text{model}} \equiv n_A M_A + n_B M_B$ with values derived from fit models. The data points correspond in color to their model. 1σ error bars have been plotted but are too small to see.

On the other hand, the SGM fits to the GMM produces a biased inferred Hubble diagram and biased inferences for the cosmological parameters. Compare Fig. 4 to the right plot of Fig. 3. Notice that the results of the two fits no longer cross near $z = 0.775$ once cosmological parameters are fit simultaneously with SN Ia population

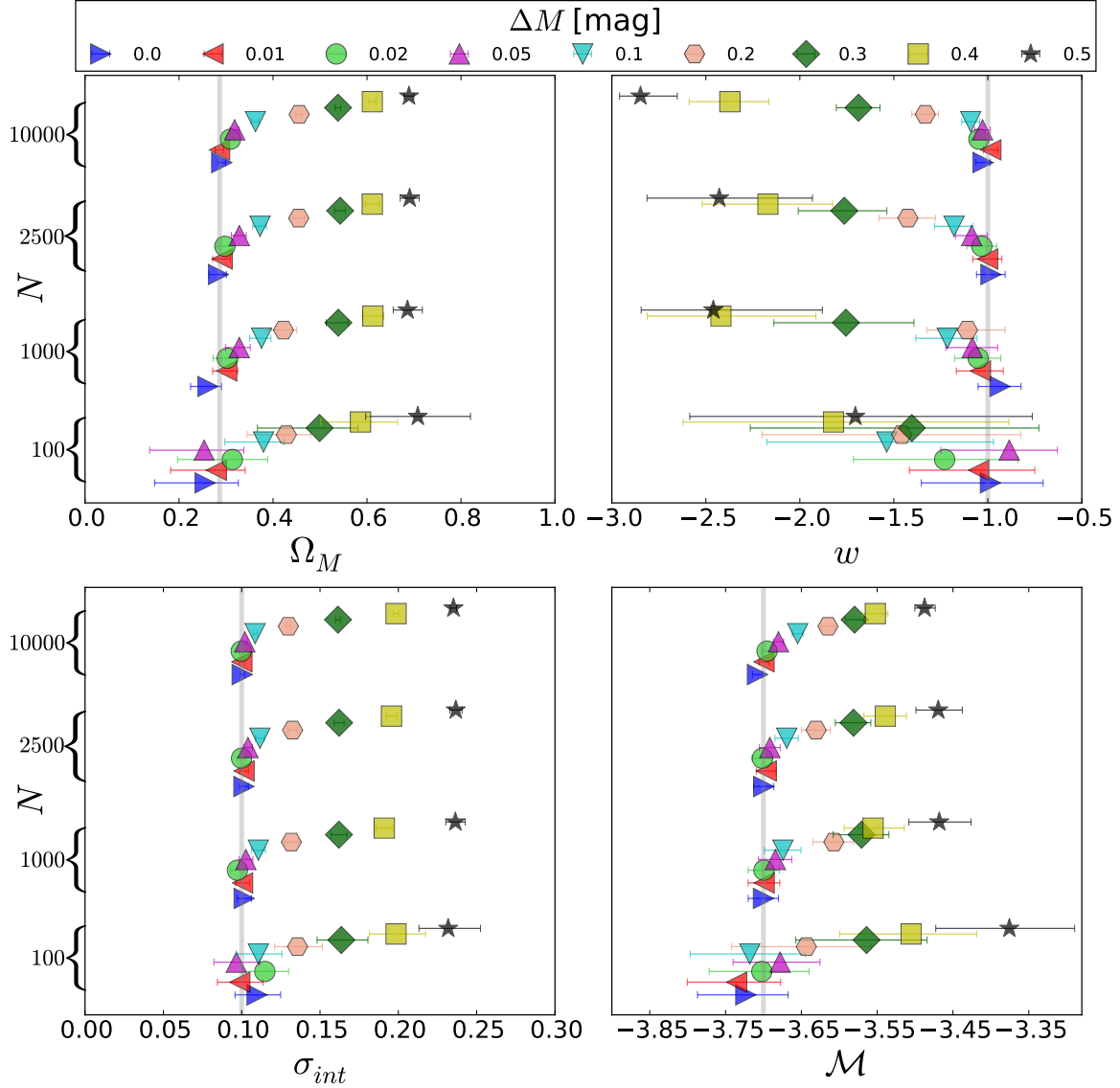


Figure 5. The median and 68% confidence region from the MCMC analysis using a SGM likelihood plotted within the limits of each prior (except \mathcal{M}). Each ΔM is given its own color and shape. To avoid overlap in the error bars, we present the increasing ΔM values with a small offset in ordinate value within each N bracket. The grey vertical line in each plot marks the fiducial value of that parameter.

parameters. The SGM fits to the GMM mock data result in cosmological parameters and SN Ia population parameters that are simultaneously significantly biased. As a result, the inferred Hubble diagram differs from the true underlying dependence of distance modulus on redshift. Most importantly, the bias in the cosmological parameters is significant. We infer $\Omega_M = 0.69 \pm 0.01$ and $w = -2.85^{+0.19}_{-0.11}$ and rule out the true underlying cosmology with high confidence. Of course, this model with $\Delta M = 0.5$ mag is extreme, but we will now move on to a discussion of inferred cosmological parameters in each of our 108 mock data sets and show that viable two-population SN Ia models yield biases in cosmology that are non-negligible compared to statistical errors.

We present medians and 68% confidence regions of the fitted parameters by combining the MCMC results from the 3 different data sets at each value of N and at each value of ΔM . We define the 68% confidence region as the area contained within the 16th and 84th percentiles,

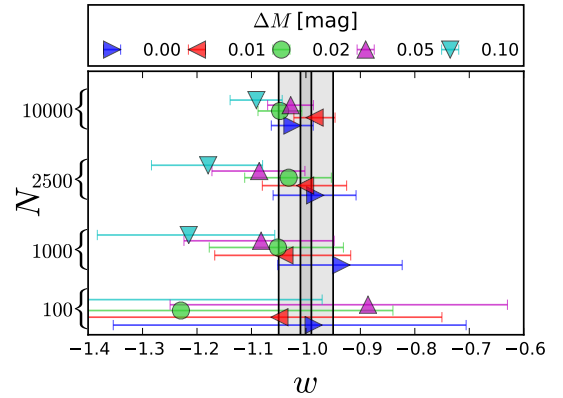


Figure 6. A close up view of the Top Right panel from Figure 5 to show the induced bias on w . We here only focus on the empirically plausible values $0.0 < \Delta M < 0.1$ mag. The vertical lines indicate ranges of ± 0.05 and ± 0.01 in w for reference.

which enforces an equal probability in the tails at either end of the posterior distribution. In order to combine the three data sets, we calculate the average of the medians, and we calculate the 16th and 84th percentiles as

$$\sigma_{\%} = \sqrt{\frac{\sigma_{\%,1}^2 + \sigma_{\%,2}^2 + \sigma_{\%,3}^2}{3}}, \quad (9)$$

where $\sigma_{\%}$ is the 16th or 84th percentile and $\sigma_{\%,i}$ corresponds to the 16th or 84th percentile calculated from the i^{th} data set.

Fig. 5 shows the medians and 68% confidence regions in the inferred parameters in our fits using a SGM to describe GMM mock data. As Fig. 5 clearly shows, for $\Delta M = 0$, the inferred parameters are unbiased: the true, underlying value of each of the cosmological parameters is inferred to within statistical precision. This is unsurprising. We have assumed that both sub-populations have the same intrinsic dispersion, so a model in which $\Delta M = 0$ is tantamount to a SGM for SNeIa. This is nothing more than a validation of this procedure for a single population of SNeIa. Models with $\Delta M \neq 0$ correspond to GMM models. Both cosmological and SN Ia population model parameters exhibit increasing biases as ΔM increases. Moreover, many of these biases are quite statistically significant suggesting that it is possible to rule out the correct underlying models due to these systematic errors. We note that in some cases ($\Delta M \gtrsim 0.4$) the inferred values of w are strongly influenced by the hard prior $w > -3$ that we have enforced. Table 2 summarizes only the results for cosmological parameters with 68% confidence regions for SGM and GMM results.

Fig. 6 is an analogous plot focusing on the inferred values of w , which is the primary science goal of dark energy probes, and observationally-plausible values of $\Delta M \leq 0.1$. Even in this restricted range of ΔM it is apparent that neglecting the possibility of multiple populations can lead to biases in the inferred value of w that are non-negligible compared to the statistical errors in these parameters. This is clearly a challenge to precision measurements of the dark energy equation of state that must be overcome in order to fully exploit SNeIa.

In comparison, the inferred parameters in the GMM model fits to the GMM mock data can be seen in Figure 7 and Figure 8. In *all* such cases we recover the correct cosmological parameters to within statistical precision. Indeed this is not entirely surprising because this is now a fit with a model that correctly describes the mock data. Indeed, we are able to infer all of the model parameters in an unbiased way except for $n'_{A,0}$ and $n_{A,0}$ when $\Delta M \lesssim 0.2$ mag. The fiducial values are recovered within the 99% confidence region for the intercept $n_{A,0}$ and within $\sim 99.9\%$ confidence region for the slope $n'_{A,0}$. It is clear that $n'_{A,0}$ and $n_{A,0}$ are biased in Figure 7 in a way that favors less redshift dependence (smaller $n'_{A,0}$) except for large shifts in peaks of the two populations. Even though these parameters are biased, they do not introduce an increase in the variance of cosmological parameters. This counter intuitive result can be explained through Figure 10, which shows that the posterior distributions of the population versus cosmological parameters are parallel to the population parameters meaning they have little to no degeneracy with cosmological pa-

rameters. When ΔM is sufficiently small, data with the precision and size of our mock data sets cannot clearly distinguish the two peaks because the separation between the peaks is comparable to the dispersion in any one of the sub-populations. It is important to note that cosmological parameters can be strongly biased despite the fact that a fit to the underlying data *cannot* clearly distinguish the two populations. This is relevant to the results of the following subsection.

Clearly, an underlying model in which $\Delta M = 0$ and $\sigma_A = \sigma_B$ can be described by a SGM with no bias. Using a GMM model to describe such data introduces additional parameters and necessarily leads to less restrictive constraints on the cosmological parameters of interest. This loss in precision is the cost of using a model with the parameter freedom to account for the possibility of multiple SNeIa sub-populations. For a data set with the precision expected of $N = 2500$ ($N = 10,000$) SNeIa, the loss of precision in Ω_M is $\sim 20\%$ ($\sim 25\%$) while the loss of precision in w is approximately $\lesssim 1\%$ ($\lesssim 3\%$). This very moderate cost in precision greatly outweighs the potential $\sim 2\sigma$ statistical error that can be induced by treating a two populations of SNeIa with $\Delta M \sim 0.1$ as a single populations (see Table 2). This finding reaffirms that the precision does not significantly decrease when these population parameters are added to the model.

Figure 9 shows the cosmological parameter posteriors from one data set for the interesting case of $N = 1,000$ and $\Delta M = 0.1$ mag. These numbers are interesting since the JLA has ~ 1000 SNeIa, and the current estimated discrepancy in Hubble residuals is equivalent to $\Delta M \sim 0.1$ mag. The contours continue to show that the GMM is less biased but also slightly less precise. These are not large offsets, but it could lead to a small systematic error in the next stages of observational cosmology.

5.3. Model Selection

To determine if the additional complexity of a given model is demanded by the (mock) data, we use the information criteria described in Section 4.2. In order to compare two models, one can compute the information criteria for each and take the difference between the two results. For example, if we compute the AIC for each model, we would compute $\Delta\text{AIC} = \text{AIC}_{\text{GMM}} - \text{AIC}_{\text{SGM}}$ where AIC_{GMM} is the value of AIC in the GMM model and likewise for AIC_{SGM} . We follow this convention, subtracting the SGM criteria from the GMM criteria, so that lower values of the difference between information criteria (IC) favor the GMM model. With these conventions, any change in information criteria (generically, ΔIC) will favor the GMM if $\Delta\text{IC} < 0$ and strongly favor the GMM if $\Delta\text{IC} < -5$. Conversely, a positive ΔIC favors the SGM while $\Delta\text{IC} > 5$ strongly favors the SGM.

We look for the minimum ΔM for each N that strongly favors the GMM. The results of this comparison are summarized Table 3 for all of the IC and in Fig. 11 for the AIC alone. The AIC, BIC, and DIC all give very comparable results. Notice that ΔM must be relatively large in order for the IC to indicate that the data demand a two-population model of SNeIa. Indeed, a data set of $N \gtrsim 10,000$ SNeIa is required in order for the IC to prefer strongly the GMM with $\Delta M \sim 0.1$ over the SGM.

Table 2
The median and 68% confidence region of Ω_M and w for all N and ΔM .

N	Model	Ω	$\Delta M = 0.0$	0.01	0.02	0.05	0.1	0.2	0.3	0.4	0.5
100	GMM	Ω_M	$0.234^{+0.071}_{-0.099}$	$0.265^{+0.061}_{-0.093}$	$0.295^{+0.075}_{-0.104}$	$0.23^{+0.083}_{-0.106}$	$0.356^{+0.049}_{-0.087}$	$0.408^{+0.056}_{-0.088}$	$0.406^{+0.122}_{-0.138}$	$0.428^{+0.112}_{-0.126}$	$0.303^{+0.089}_{-0.089}$
100	SGM	Ω_M	$0.255^{+0.071}_{-0.107}$	$0.279^{+0.062}_{-0.097}$	$0.314^{+0.075}_{-0.117}$	$0.253^{+0.085}_{-0.116}$	$0.38^{+0.044}_{-0.082}$	$0.428^{+0.083}_{-0.084}$	$0.499^{+0.082}_{-0.132}$	$0.586^{+0.079}_{-0.084}$	$0.708^{+0.112}_{-0.111}$
100	GMM	w	$-0.988^{+0.265}_{-0.346}$	$-1.047^{+0.286}_{-0.357}$	$-1.189^{+0.352}_{-0.470}$	$-0.875^{+0.231}_{-0.331}$	$-1.46^{+0.513}_{-0.627}$	$-1.471^{+0.615}_{-0.721}$	$-1.265^{+0.513}_{-0.758}$	$-1.447^{+0.626}_{-0.740}$	$-1.341^{+0.466}_{-0.554}$
100	SGM	w	$-0.986^{+0.28}_{-0.368}$	$-1.049^{+0.298}_{-0.369}$	$-1.23^{+0.389}_{-0.485}$	$-0.886^{+0.256}_{-0.363}$	$-1.538^{+0.567}_{-0.638}$	$-1.459^{+0.634}_{-0.742}$	$-1.404^{+0.676}_{-0.859}$	$-1.821^{+0.933}_{-0.800}$	$-1.705^{+0.941}_{-0.881}$
1000	GMM	Ω_M	$0.247^{+0.032}_{-0.040}$	$0.282^{+0.030}_{-0.038}$	$0.285^{+0.028}_{-0.036}$	$0.314^{+0.027}_{-0.035}$	$0.363^{+0.025}_{-0.035}$	$0.361^{+0.052}_{-0.061}$	$0.289^{+0.040}_{-0.048}$	$0.287^{+0.036}_{-0.043}$	$0.295^{+0.033}_{-0.041}$
1000	SGM	Ω_M	$0.261^{+0.029}_{-0.037}$	$0.301^{+0.024}_{-0.030}$	$0.303^{+0.024}_{-0.030}$	$0.328^{+0.023}_{-0.029}$	$0.376^{+0.020}_{-0.025}$	$0.422^{+0.028}_{-0.039}$	$0.539^{+0.021}_{-0.026}$	$0.612^{+0.023}_{-0.023}$	$0.686^{+0.032}_{-0.030}$
1000	GMM	w	$-0.92^{+0.108}_{-0.113}$	$-1.014^{+0.117}_{-0.125}$	$-1.025^{+0.116}_{-0.122}$	$-1.06^{+0.129}_{-0.139}$	$-1.184^{+0.155}_{-0.167}$	$-0.994^{+0.173}_{-0.197}$	$-0.993^{+0.152}_{-0.170}$	$-1.003^{+0.146}_{-0.162}$	$-1.014^{+0.140}_{-0.150}$
1000	SGM	w	$-0.935^{+0.112}_{-0.117}$	$-1.039^{+0.122}_{-0.128}$	$-1.051^{+0.120}_{-0.126}$	$-1.082^{+0.134}_{-0.142}$	$-1.215^{+0.158}_{-0.168}$	$-1.108^{+0.200}_{-0.217}$	$-1.755^{+0.362}_{-0.383}$	$-2.42^{+0.505}_{-0.392}$	$-2.46^{+0.580}_{-0.384}$
2500	GMM	Ω_M	$0.267^{+0.024}_{-0.028}$	$0.281^{+0.020}_{-0.024}$	$0.285^{+0.020}_{-0.025}$	$0.311^{+0.021}_{-0.027}$	$0.36^{+0.017}_{-0.023}$	$0.318^{+0.032}_{-0.029}$	$0.296^{+0.024}_{-0.027}$	$0.298^{+0.024}_{-0.028}$	$0.239^{+0.029}_{-0.032}$
2500	SGM	Ω_M	$0.284^{+0.017}_{-0.020}$	$0.292^{+0.017}_{-0.020}$	$0.298^{+0.016}_{-0.019}$	$0.328^{+0.015}_{-0.017}$	$0.372^{+0.013}_{-0.016}$	$0.455^{+0.012}_{-0.013}$	$0.543^{+0.012}_{-0.012}$	$0.611^{+0.015}_{-0.015}$	$0.691^{+0.020}_{-0.020}$
2500	GMM	w	$-0.967^{+0.073}_{-0.077}$	$-0.988^{+0.075}_{-0.078}$	$-1.012^{+0.077}_{-0.080}$	$-1.057^{+0.083}_{-0.087}$	$-1.16^{+0.100}_{-0.104}$	$-1.095^{+0.100}_{-0.116}$	$-1.02^{+0.094}_{-0.102}$	$-0.994^{+0.098}_{-0.103}$	$-0.865^{+0.083}_{-0.091}$
2500	SGM	w	$-0.983^{+0.075}_{-0.077}$	$-1.001^{+0.076}_{-0.079}$	$-1.032^{+0.078}_{-0.081}$	$-1.086^{+0.084}_{-0.087}$	$-1.18^{+0.100}_{-0.104}$	$-1.425^{+0.146}_{-0.154}$	$-1.763^{+0.226}_{-0.245}$	$-2.17^{+0.345}_{-0.349}$	$-2.429^{+0.496}_{-0.385}$
10000	GMM	Ω_M	$0.285^{+0.010}_{-0.013}$	$0.28^{+0.011}_{-0.014}$	$0.303^{+0.010}_{-0.013}$	$0.312^{+0.011}_{-0.015}$	$0.321^{+0.033}_{-0.031}$	$0.307^{+0.015}_{-0.014}$	$0.28^{+0.013}_{-0.014}$	$0.286^{+0.012}_{-0.013}$	$0.283^{+0.012}_{-0.012}$
10000	SGM	Ω_M	$0.291^{+0.009}_{-0.009}$	$0.286^{+0.009}_{-0.009}$	$0.309^{+0.008}_{-0.009}$	$0.318^{+0.009}_{-0.009}$	$0.363^{+0.008}_{-0.009}$	$0.456^{+0.007}_{-0.007}$	$0.538^{+0.006}_{-0.006}$	$0.611^{+0.008}_{-0.008}$	$0.689^{+0.010}_{-0.010}$
10000	GMM	w	$-1.016^{+0.039}_{-0.040}$	$-0.976^{+0.038}_{-0.039}$	$-1.037^{+0.041}_{-0.042}$	$-1.017^{+0.044}_{-0.044}$	$-1.028^{+0.055}_{-0.061}$	$-1.028^{+0.047}_{-0.050}$	$-0.98^{+0.046}_{-0.046}$	$-1.014^{+0.048}_{-0.050}$	$-0.991^{+0.045}_{-0.048}$
10000	SGM	w	$-1.025^{+0.038}_{-0.039}$	$-0.984^{+0.038}_{-0.038}$	$-1.047^{+0.040}_{-0.041}$	$-1.028^{+0.042}_{-0.042}$	$-1.091^{+0.047}_{-0.048}$	$-1.333^{+0.070}_{-0.072}$	$-1.688^{+0.114}_{-0.118}$	$-2.372^{+0.207}_{-0.217}$	$-2.848^{+0.196}_{-0.111}$

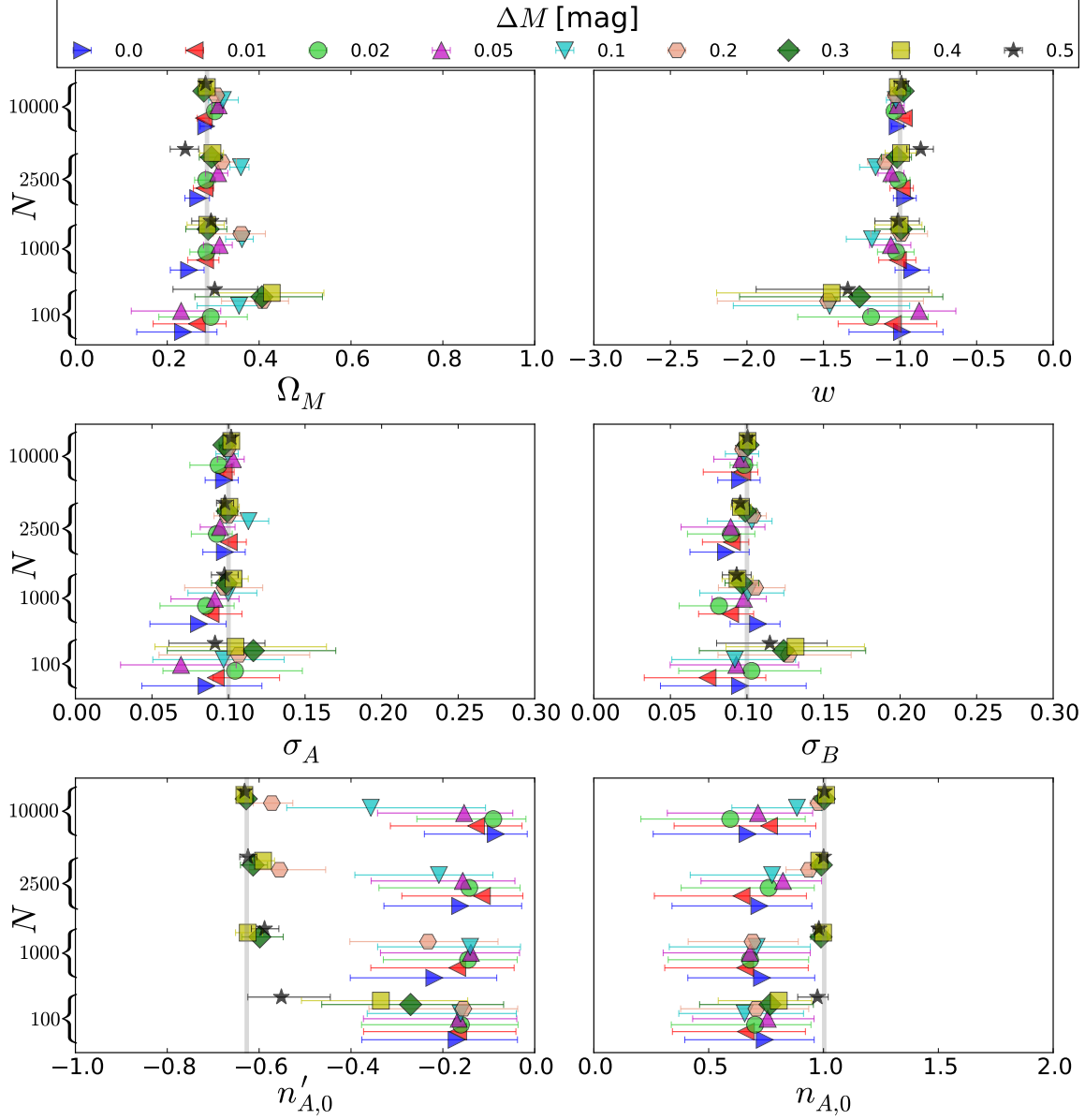


Figure 7. The GMM fit results for the same simulations as in Figure 5 and displayed with the same ordinate offsets. The GMM model correctly recovers the fiducial cosmology and accounts for the multiple input populations. The multiple-population parameters (σ_A , σ_B) are not well constrained for small ΔM , and the normalization factors ($n_{A,0}$ and $n'_{A,0}$) are even clearly biased at low ΔM due to the reduced leverage they have on the output. But the resulting cosmological parameters are well-constrained when marginalizing over the multiple-population parameters.

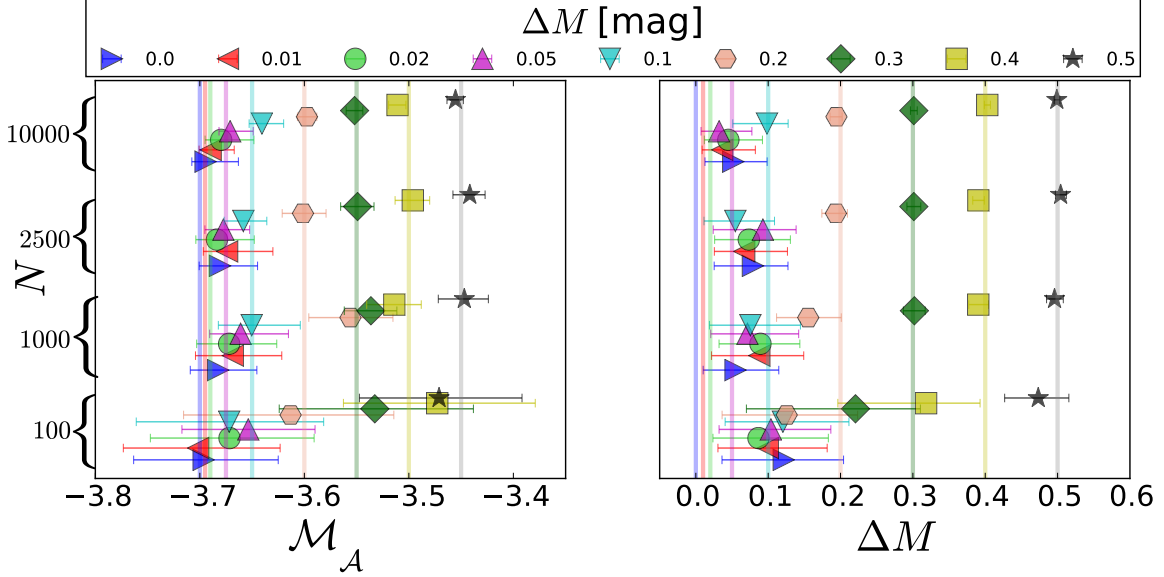


Figure 8. The GMM fit results for parameters dependent on ΔM from the same simulations as in Figures 5 and 7 and displayed with the same ordinate offsets. Each specific ΔM has a vertical line denoting the different fiducial values. While Figure 5 and Figure 7 are plotted over the entire range of the uniform prior, in this figure we focus on a range much smaller than the prior to show the detailed effect.

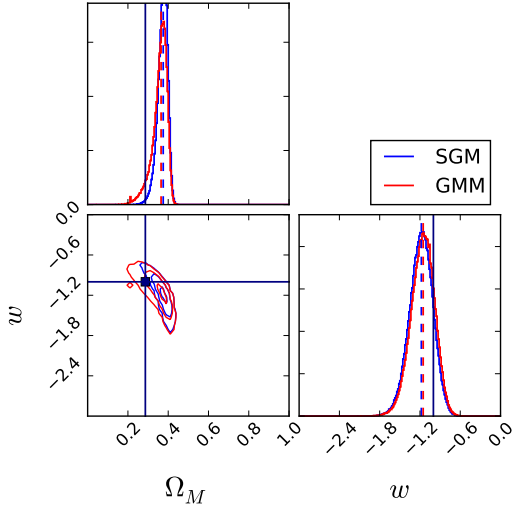


Figure 9. Cosmological contours for $\Delta M = 0.1$ mag and $N = 1000$. Made with `triangle.py` from Foreman-Mackey et al. (2014). The blue contours and histograms correspond to the mock data being fit with SGM likelihood and the red contours and histograms correspond to the GMM likelihood. The dashed lines are the medians of the populations. The dark navy lines are the fiducial values. The GMM is less precise but also less biased.

Table 3
Minimum ΔM (in mag)
with strong evidence for
GMM.

N	AIC	BIC	DIC
100	0.40	0.45	0.41
1000	0.21	0.25	0.23
2500	0.12	0.21	0.23
10000	0.10	0.14	0.10

There is an important point regarding the interpretation of the results of this section in conjunction with those of the previous subsections. The fact that the data may not *demand* a GMM to describe SNeIa does *not* mean that a multiple-population SNeIa model is not *necessary*. As we have shown, statistically significant biases in cosmological parameters can be inferred when two-population data are analyzed as a single population, even when the information criteria do not unambiguously demand the GMM rather than the SGM. If by “necessary” one means that the model is needed in order to infer unbiased cosmological parameters, then the GMM may be necessary even when the IC yield only marginal evidence. IC that do not clearly demand the more complex model (the GMM in this case) are not sufficient justification for using only the simpler model (the SGM in this case) in cosmological analyses because significant parameter biases may still be realized using the simpler approach.

6. DISCUSSION

6.1. Usage of the SGM

The SGM was meant to be representative of the latest supernova cosmology analysis, namely the Joint Lightcurve Analysis (JLA); however, the SGM cannot be directly compared to the JLA. Unlike the SGM, the JLA further standardizes each supernova by applying an offset to the absolute magnitude of each supernova using an empirically-derived step function in host galaxy mass. This standardization follows from the observed Hubble residual trend with host galaxy properties that was one of the motivations for introducing multiple populations. Leaving out the host galaxy standardization enables this present study to avoid any unintended bias from using the step function, conceptually compare the SGM to the GMM, and create a general framework that can be applied to other systematics. The goal of this paper is not to implement a new model for the correlation between

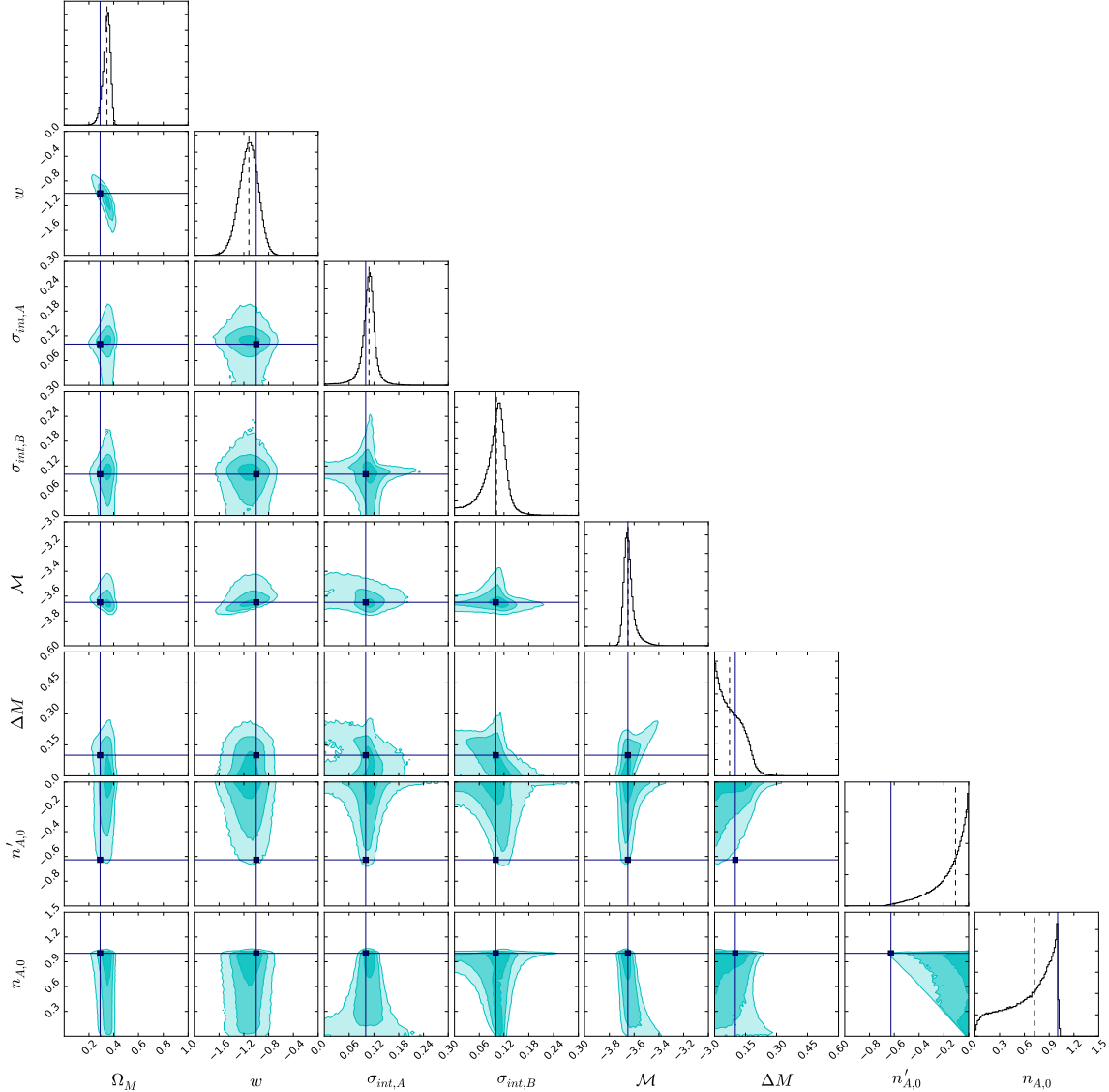


Figure 10. Full triangle plot of posterior distributions from a GMM likelihood for a single data set with $N = 1,000$ and $\Delta M = 0.1$ mag. Made with `triangle.py` from Foreman-Mackey et al. (2014)

the SN Hubble residual and host-galaxy properties, but to introduce a statistical framework in which to implement a future model.

The likelihoods for the SGM and JLA are the same except that the JLA utilizes the host galaxy mass standardization and a full covariance matrix. JLA uses a χ^2 minimization for parameter estimation, which is equivalent to maximizing a Gaussian likelihood. The JLA uses a frequentist approach with χ^2 minimization, but we use the SGM to explore parameter space through Bayesian statistics with MCMC. However, χ^2 minimization and the SGM likelihood analysis both use a Gaussian single-point estimate of the SN corrected brightness to infer cosmological parameters. Using single-point estimate (μ, σ) does not provide framework to deal with insufficient population modeling and data with large error bars on parameters used for systematics modeling⁵, both of which are found in the current data sets. This present paper

⁵ The mass of each host galaxy is determined from photometry

shows that updating the likelihood to incorporate non-Gaussian effects can remove bias on cosmology without precise modeling of the underlying populations.

6.2. Connection To Astrophysical Properties

A relationship with host galaxy mass is currently used to correct SN Ia apparent brightness; however, host mass must be an indicator of a different galactic property that has a connection to the brightness of a supernova such as local metallicity, star formation rate, and stellar population age (Johansson et al. 2013). One possible explanation for the host mass effect is different progenitor ages. The overall mass of the galaxy is correlated with progenitor age through stellar population ages. SNeIa occur in both active and passive star forming regions, which implies that they have both short delay times ($\sim 100 - 500$ Myr) and long delay times (~ 5 Gyr) between progenitor formation and supernova event (Mannucci et al. 2005;

in the JLA sample has a typical uncertainty of ~ 0.8 dex.

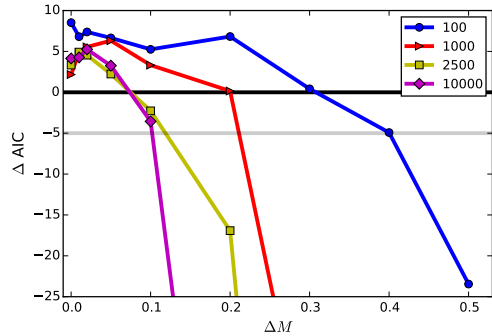


Figure 11. $AIC(GMM) - AIC(SGM)$ for $N=[100,1000,2500,10000]$ as a function of the separation of peaks. GMM is considered strongly favored once $\Delta AIC < -5$.

Scannapieco & Bildsten 2005; Mannucci et al. 2006; Sullivan et al. 2006). The different progenitor ages could be motivated by different channels for a thermonuclear explosion: single degenerate (SD) where a white dwarf accretes matter from a main sequence or red giant companion (Whelan & Iben 1973) and double degenerate (DD) where two white dwarfs merge (Webbink 1984). The SD and DD can both explain the population with short delay times; however, SD models do not support the long delay times (Greggio (2005), see Maoz et al. (2014) for comprehensive review).

Several papers have begun to examine the connections between host galaxy mass and stellar population ages. Johansson et al. (2013) showed that the stretch-host galaxy mass relationship is caused by the correlation between host galaxy mass and stellar population age. Childress et al. (2013) fit the Hubble residual versus host galaxy mass with different functional forms and examined different physical causes of the relationship. They found the best physical link to the step function was the evolution of the prompt fraction of SN Ia progenitors, but the fit is not adequate enough to be the only source of the effect. Childress et al. (2014) focused on modeling stellar population age as a function of host galaxy mass at different redshifts. The paper showed a bimodal distribution in progenitor age versus stellar mass and that this bimodality is evident out to a redshift $\lesssim 0.5$. These results clearly motivate adopting a GMM approach where the two populations changing with redshift. Unfortunately, the way the populations evolve with redshift is determined through star formation histories and delay time distributions, which is considerably more complicated than the simple linear evolution probed here. Creating better astrophysical models for the evolution of systematics is an active area of research, and we present this generalized PDF approach as the appropriate framework to incorporate them into.

7. CONCLUSION

This paper explored expanding supernovae analyses into a broader scope with a generalized likelihood model. For illustration we used a toy example of two-population GMM with a simple linear evolution in relative population with redshift. We explored different distributions of likelihood functions and showed that in mock data sets using our toy GMM example multiple SNeIa sub-populations may lead to significant biases in cosmologi-

cal parameters inferred from SNeIa data. In particular, when $N = 1,000$ and $\Delta M = 0.1$ mag, biases may be 2-4 times that of the statistical uncertainty. Incorporate this model into the PDF removes systematic errors (biases) in inferred cosmological parameters at a small statistical cost, roughly 2% in the marginalized uncertainty on w . Large data sets ($N > 10,000$) are necessary to yield unambiguous evidence of multiple populations according to various model selection criteria. However, even when model selection does not clearly favor multiple populations, the presence of multiple populations in the data can severely bias cosmological parameters. Our approach of modeling the possibility of multiple populations not only mitigates biases from them, but also yields a small penalty in precision if there is only one population.

The existence of multiple populations is still being debated as seen in Jones et al. (2015), which advocates for a single population; however, a GMM likelihood has the capability of determining if there is only one population and thus is a more rigorous way to analyze the data to ensure more systematics are included.

We have assumed an example model of two populations with a difference in the absolute magnitude, but there are clearly other channels in which separate populations might be expressed depending on the astrophysical cause. It is possible that a different supernova property can better parameterize the stellar population age of the progenitor. If we did not use the width-color-corrected apparent magnitude, then the apparent magnitudes would be defined as $M_X \equiv M_{Bband,X} - \alpha x_1 + \beta \mathcal{C}$, where x_1 is the stretch calculated from each supernova light curve, α is the stretch parameter determined for the entire supernova population, \mathcal{C} is the color of each supernova at time of maximum light, and β is the color parameter determined for the entire supernova population. One example has been provided by Milne et al. (2015) which shows two different populations with a difference in near ultraviolet (NUV) $u - v$ color of 0.4 magnitudes (0.1 mag in $b - v$) with the relative fractions of populations evolving with redshift. This color dependence would fit nicely into our framework since we could alternatively model the absolute magnitude as $M_X \equiv M_{Bband} - \alpha x_1 + \beta_X \mathcal{C}$.

This framework is tested with the host galaxy mass dependence as an example; however, it is suitable for accounting for any systematic that may have multiple values based on supernova parameters. For example, surveys with different selection effects could also be included as different PDFs, either in intrinsic distribution or in redshift evolution, for each survey. Corrections for Malmquist bias (Malmquist 1936) could be handled more cleanly by using the full PDF instead of using the mean computed correction for the sample (e.g., Perrett et al. 2010; Conley et al. 2011; Rest et al. 2014; Scolnic et al. 2014; Betoule et al. 2014) or priors on the light-curve fitting parameters applied on per-object basis (e.g., Wood-Vasey et al. 2007). Currently forward-modeling approaches that simulate entire surveys (e.g., SNANA Kessler et al. 2009, 2010) carry through this modeling all the way; we believe there can be significant gains in translating much of this information into empirical PDFs that can then be interpolated and used in a generalized full-likelihood fitting (work towards this has begun in Rubin et al. (2015)).

Supernova cosmology would benefit from incorporating a non-Gaussian likelihood with an MCMC analysis to model the many systematics involved in order to remove biases with a minimal precision loss.

We thank the referee for constructive comments, which have improved this paper. We thank Phil Marshall and Carles Badenes for useful discussions. KAP was supported in this work by NSF AST-1028162, and by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-AC05-06OR23100. M.W.-V. was supported in part by DOE de-sc0011834 and through sabbatical support from Stanford/SLAC. The work of ARZ is supported in part by the DOE through grant de-sc0007914 and by the Pittsburgh Particle physics, Astrophysics, and Cosmology Center (Pitt PACC) at the University of Pittsburgh.

REFERENCES

- Akaike, H. 1974, *IEEE Transactions on Automatic Control*, 19, 716
- Astier, P., Balland, C., Brescia, M., et al. 2014, *A&A*, 572, A80
- Baade, W. 1938, *ApJ*, 88, 285
- Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A22
- Box, G. E. P., & Jenkins, G. M., eds. 1976, *Time series analysis. Forecasting and control*
- Cao, Y., Kulkarni, S. R., Howell, D. A., et al. 2015, *Nature*, 521, 328
- Childress, M., Aldering, G., Antilogus, P., et al. 2013, *ApJ*, 770, 108
- Childress, M. J., Wolf, C., & Zahid, H. J. 2014, *MNRAS*, 445, 1898
- Conley, A., Guy, J., Sullivan, M., et al. 2011, *ApJS*, 192, 1
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Foreman-Mackey, D., Price-Whelan, A., Ryan, G., et al. 2014, *triangle.py v0.1.1*, doi:10.5281/zenodo.11020
- Gelman, A., Hwang, J., & Vehtari, A. 2014, *Statistics and computing*, 24
- Goodman, J., & Weare, J. 2010, *Comm. App. Math. and Comp. Sci*, 5
- Greggio, L. 2005, *A&A*, 441, 1055
- Greggio, L., Renzini, A., & Daddi, E. 2008, *MNRAS*, 388, 829
- Gupta, R. R., D’Andrea, C. B., Sako, M., et al. 2011, *ApJ*, 740, 92
- Helbig, P. 2015, *MNRAS*, 453, 3975
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, 208, 19
- Johansson, J., Thomas, D., Pforr, J., et al. 2013, *MNRAS*, 435, 1680
- Jones, D. O., Riess, A. G., & Scolnic, D. M. 2015, *ArXiv e-prints*, arXiv:1506.02637
- Kelly, P. L., Filippenko, A. V., Burke, D. L., et al. 2015, *Science*, 347, 1459
- Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, *ApJ*, 715, 743
- Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, *PASP*, 121, 1028
- . 2010, *Astrophysics Source Code Library*, ascl:1010.027
- Kowal, C. T. 1968, *AJ*, 73, 1021
- Lampeitl, H., Smith, M., Nichol, R. C., et al. 2010, *ApJ*, 722, 566
- Liddle, A. R. 2007, *MNRAS*, 377, L74
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, *ArXiv e-prints*, arXiv:0912.0201
- Malmquist, K. G. 1936, *Stockholms Observatoriums Annaler*, 12, 7
- Mannucci, F., Della Valle, M., & Panagia, N. 2006, *MNRAS*, 370, 773
- Mannucci, F., Della Valle, M., Panagia, N., et al. 2005, *A&A*, 433, 807
- Maoz, D., Mannucci, F., & Nelemans, G. 2014, *ARA&A*, 52, 107
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *J. Chem. Phys.*, 21, 1087
- Milne, P. A., Foley, R. J., Brown, P. J., & Narayan, G. 2015, *ApJ*, 803, 20
- Olling, R. P., Mushotzky, R., Shaya, E. J., et al. 2015, *Nature*, 521, 332
- Pearson, K. 1894, *Phil. Trans. Roy. Soc. London*, A, 71
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, 517, 565
- Perrett, K., Balam, D., Sullivan, M., et al. 2010, *AJ*, 140, 518
- Phillips, M. M. 1993, *ApJ*, 413, L105
- Rest, A., Scolnic, D., Foley, R. J., et al. 2014, *ApJ*, 795, 44
- Riess, A. G., Press, W. H., & Kirshner, R. P. 1996, *ApJ*, 473, 88
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, 116, 1009
- Rigault, M., Copin, Y., Aldering, G., et al. 2013, *A&A*, 560, A66
- Rigault, M., Aldering, G., Kowalski, M., et al. 2015, *ApJ*, 802, 20
- Rubin, D., Aldering, G., Barbary, K., et al. 2015, *ApJ*, 813, 137
- Scannapieco, E., & Bildsten, L. 2005, *ApJ*, 629, L85
- Schwarz, G. 1978, *Ann. Statist.*, 6, 461
- Scolnic, D., Rest, A., Riess, A., et al. 2014, *ApJ*, 795, 45
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, *ArXiv e-prints*, arXiv:1503.03757
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. 2002, *Journal of the Royal Statistical Society*, 64, 583
- Sullivan, M., Le Borgne, D., Pritchett, C. J., et al. 2006, *ApJ*, 648, 868
- Sullivan, M., Conley, A., Howell, D. A., et al. 2010, *MNRAS*, 406, 782
- Tripp, R. 1998, *A&A*, 331, 815
- Webbink, R. F. 1984, *ApJ*, 277, 355
- Whelan, J., & Iben, Jr., I. 1973, *ApJ*, 186, 1007
- Wood-Vasey, W. M., Miknaitis, G., Stubbs, C. W., et al. 2007, *ApJ*, 666, 694
- Zentner, A. R., & Bhattacharya, S. 2009, *ApJ*, 693, 1543